



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G06F 17/30	A1	(11) International Publication Number: WO 00/08572 (43) International Publication Date: 17 February 2000 (17.02.00)		
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top; padding: 5px;"> (21) International Application Number: PCT/US99/17359 (22) International Filing Date: 2 August 1999 (02.08.99) (30) Priority Data: 09/130,420 6 August 1998 (06.08.98) US (71) Applicant: GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES, LLC [US/US]; 236 Huntington Avenue, Boston, MA 02115-4701 (US). (72) Inventors: SCHABES, Yves; 770 Boylston Street #12B, Boston, MA 02199 (US). ROCHE, Emmanuel; 64 Livernmore Street, Belmont, MA 02178 (US). BROWN, Ryan; 3764 Lincoln Road, Santa Barbara, CA 93110 (US). (74) Agent: NUGENT, Elizabeth, E.; Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109 (US). </td> <td style="width: 50%; vertical-align: top; padding: 5px;"> (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> </td> </tr> </table>			(21) International Application Number: PCT/US99/17359 (22) International Filing Date: 2 August 1999 (02.08.99) (30) Priority Data: 09/130,420 6 August 1998 (06.08.98) US (71) Applicant: GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES, LLC [US/US]; 236 Huntington Avenue, Boston, MA 02115-4701 (US). (72) Inventors: SCHABES, Yves; 770 Boylston Street #12B, Boston, MA 02199 (US). ROCHE, Emmanuel; 64 Livernmore Street, Belmont, MA 02178 (US). BROWN, Ryan; 3764 Lincoln Road, Santa Barbara, CA 93110 (US). (74) Agent: NUGENT, Elizabeth, E.; Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109 (US).	(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(21) International Application Number: PCT/US99/17359 (22) International Filing Date: 2 August 1999 (02.08.99) (30) Priority Data: 09/130,420 6 August 1998 (06.08.98) US (71) Applicant: GLOBAL INFORMATION RESEARCH AND TECHNOLOGIES, LLC [US/US]; 236 Huntington Avenue, Boston, MA 02115-4701 (US). (72) Inventors: SCHABES, Yves; 770 Boylston Street #12B, Boston, MA 02199 (US). ROCHE, Emmanuel; 64 Livernmore Street, Belmont, MA 02178 (US). BROWN, Ryan; 3764 Lincoln Road, Santa Barbara, CA 93110 (US). (74) Agent: NUGENT, Elizabeth, E.; Choate, Hall & Stewart, Exchange Place, 53 State Street, Boston, MA 02109 (US).	(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>			
(54) Title: SEARCH AND INDEX HOSTING SYSTEM				
(57) Abstract <p>The system initiates a search at a first network site for user-specified data in a remote database at a second network site and conducts the search at a third network site (e.g., at a host computer's site). To begin, the system receives, at the first network site, a provider identifier associated with the database from the second network site. Thereafter, the user-specified data is input at the first network site, following which the user-specified data and the provider identifier are output from the first network site to the third network site. The system then searches for the user-specified data in a database at the third network site using the provider identifier. This database at the third network site includes data that corresponds to data stored in the remote database at the second network site.</p>				
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>The system initiates a search at a first network site for user-specified data in a remote database at a second network site and conducts the search at a third network site (e.g., at a host computer's site). To begin, the system receives, at the first network site, a provider identifier associated with the database from the second network site. Thereafter, the user-specified data is input at the first network site, following which the user-specified data and the provider identifier are output from the first network site to the third network site. The system then searches for the user-specified data in a database at the third network site using the provider identifier. This database at the third network site includes data that corresponds to data stored in the remote database at the second network site.</p> </div> <div style="width: 50%;"> <pre> graph TD START([START]) --> S401[INPUT INFORMATION IDENTIFYING THE WEB SITE] S401 --> S402[ASSIGN PROVIDER IDENTIFIER TO THE SITE] S402 --> S403[SEARCH THROUGH THE WEB SITE IN ORDER TO IDENTIFY LOCATIONS OF DOCUMENTS THEREIN] S403 --> S404[STORE LOCATIONS OF DOCUMENTS IN THE SITE WITH PROVIDER IDENTIFIER] S404 --> S405[INDEX DOCUMENTS IN THE WEB SITE] S405 --> S406[STORE PROVIDER IDENTIFIER TOGETHER WITH INDEX OF DOCUMENTS IN THE WEB SITE] S406 --> S407[STORE COMPRESSED DATA CORRESPONDING TO DATA IN THE WEB SITE] S407 --> S408[OUTPUT THE PROVIDER IDENTIFIER TO THE WEB SITE] S408 --> END([END]) </pre> </div> </div>				

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

SEARCH AND INDEX HOSTING SYSTEM

5

BACKGROUND OF THE INVENTION

Field Of The Invention

10 The present invention relates to a system for searching a remote database, and more particularly to a system which indexes documents in the database, which identifies documents in the index that include user-specified data, and which outputs a list of documents that contain such data and, optionally, excerpts from those documents. The invention has particular utility in connection with text indexing and retrieval
15 systems, such as World Wide Web search engines.

Description Of The Related Art

20 In general, search engines search through a database for specific data and retrieve titles of documents in the database which contain that data. For example, World Wide Web search engines, such as Altavista™ and Yahoo!®, provide users with the ability to search the Web for documents containing user-specified words, phrases, or the like. However, conventional search engines, and Web search engines in particular, suffer from a drawback in that they do not allow a user to direct
25 a search to a single database.

30 More specifically, conventional Web search engines, such as those noted above, operate by generating an index for all sites on the Web, and then retrieving data from that index in response to user queries. Since these search engines generate an index for all sites on the Web, however, they are limited to searching the entire Web. This is disadvantageous, particularly for those users who only want to search specific sites.

In response to the foregoing drawbacks in the art, software manufacturers have developed site-specific searching systems, such as Ultraseek™, which ostensibly allow users to limit their searches to specific Web sites. To use these systems, however, they must be installed at each Web site, e.g., by the Web site's provider. Once installed, the systems create an index of the Web site at the provider's location. Thereafter, when a user accesses the Web site and inputs a search query, code at the site searches that index for the query, and relays the results of the search back to the user.

While the foregoing types of site-specific searching systems address some of the problems associated with conventional Web search engines, such as Yahoo!® and the like, they have several drawbacks. For example, their installation and subsequent maintenance can be costly and time consuming. As a result, Web site providers often choose not to install such systems at their sites. Additional problems arise with these systems in cases where a Web site is maintained by a Web site hosting company, as opposed to by the provider itself. That is, in these cases, in addition to the above problems, problems relating to licensing and the like arise, which make it difficult to implement conventional site-specific searching systems in a cost-effective manner.

In addition to the foregoing drawbacks, conventional database searching systems provide the user with only names/titles of documents in response to a query. For example, conventional Web search engines provide only the name of a document containing a search term, together with a uniform resource locator ("URL") for that document. As a result, it is not always possible for the user to determine which of the retrieved documents is relevant without actually linking to, and opening, the document. This can slow down searching significantly.

Thus, there exists a need for a searching system which provides the user with the ability to search a portion of a database, such as one or more sites on the World Wide Web, and which is more cost effective, efficient, and easy to use than the conventional systems described above. In addition, there exists a need for a database searching system

which is able to provide a user with the context of each search term in documents retrieved as a result of the search.

SUMMARY OF THE INVENTION

5 The present invention addresses the foregoing needs by providing a way to search through a database at one network site (e.g., a Web site) using a host computer which is at another network site. By hosting the search at a separate site, the present invention facilitates site-specific searching, as described below.

10 More specifically, according to the present invention, a Web site provider, for example, is able to create a search engine for the Web site simply by accessing the present invention via the Web and entering a request for a new account. In response to this request, the invention assigns the Web site a provider identifier, and then extracts URL(s) from
15 the Web site. Thereafter, the invention "crawls" through the site in order to create an index of the site, which comprises data from the site indexed by document (e.g., Web page) and provider identifier. Once the indexing process has been completed, the site provider need simply copy a few
20 (e.g., 10) lines of code into any sites for which searching capabilities are desired.

 Following the foregoing (i.e., the setup), each time the Web site is visited, it will automatically transmit its provider identifier to the visiting user's site. In addition, the Web site will display a search line, from which the visiting user may enter queries to search the site for
25 specific data. When such a query is entered, the query, together with the provider identifier, is passed from the user's site to the host computer's site, where the actual searching takes place. Specifically, at the host computer's site, an index corresponding to the provider identifier is retrieved from memory and searched for the data specified in the user's
30 query. Thereafter, a list of documents which contain the data (including URLs in the case of the Web) is output from the host computer's site to the user's site and displayed there.

 By conducting the search at the host computer's site, rather than

at the Web site itself, the present invention reduces the difficulties involved with installing and maintaining an entire software application at the Web site. As a result, the present invention provides a way to search specified Web sites (and other types of databases as well), which is more efficient and less costly and time consuming than the conventional site-specific searching systems described above.

Thus according to one aspect, the present invention is a system (i.e., a method, an apparatus, and computer-executable process steps) for initiating a search at a first network site for user-specified data in a remote database at a second network site and for conducting the search at a third network site (e.g., at a host computer's site). To begin, the system receives, at the first network site, a provider identifier associated with the database from the second network site. Thereafter, the user-specified data is input at the first network site, following which the user-specified data and the provider identifier are output from the first network site to the third network site. The system then searches for the user-specified data in a database at the third network site using the provider identifier. In the invention, this database at the third network site includes data that corresponds to data stored in the remote database at the second network site.

According to another aspect, the present invention is a way to configure a computerized searching system (such as the searching system resident at the host computer's site described above) so that the searching system can be used to search a database. In this aspect of the invention, information identifying the database is input, a provider identifier is assigned to the database, and a search through the database is conducted using the input information in order to identify locations of documents in the database. Thereafter, the locations of the documents in the database are stored in memory together with the provider identifier, and the documents in the database are indexed. An index of the documents is then stored in memory together with the provider identifier; and data corresponding to data in the database is also stored in memory together with the provider identifier. The provider identifier is then output to the database. As noted

above, this provider identifier is transmitted to those who visit the site.

According to still another aspect, the present invention is a system for identifying which documents in a database contain user-specified data. The system stores, in memory, indices of data in plural databases, such as those noted above. The system then receives the user-specified data and a provider identifier which corresponds to one of the plural databases, and retrieves, from memory, an index of data for a database that corresponds to the provider identified. Thereafter documents in the retrieved index that contain the user-specified data are identified, and identities thereof are output to the user.

The present invention also provides an optional feature for displaying excerpts from documents identified by a database search. In this aspect of the invention, the index of documents in the database is stored, and pointers to data segments in the database are generated based on the index. These data segments comprise target data together with data surrounding the target data. That is, assuming that the target data comprises a word which matches an input user query, the data segment for that word might comprise, e.g., five words to the left of the word, the word itself, and five words to the right of the word. When the invention searches the index for the word, it compiles a list of pointers to data segments which include the word. These data segments may then be extracted and passed to a user's site for display along with the list of documents.

Thus, according to this aspect, the invention is a system for retrieving a list of documents in a database which include user-specified data, and of retrieving one or more data segments from each document on the list. The system includes storing an index of documents from the database, the index including pointers corresponding to data in the database, where the pointers define data segments having a predetermined size. One or more documents in the database that contain the user-specified data are then identified based on the index; and a list is created which includes one or more pointers corresponding to each occurrence of the user-specified data in the identified documents. Each data segment in

the database that contains the user-specified data is extracted based on the list of pointers created in the creating step, whereafter a list of the documents in the database that contain the user-specified data is output, together with the extracted data segments.

5 By virtue of the foregoing, the invention makes it possible to display data excerpts (i.e., segments) from each document found in the search. A user may then refer to these excerpts in order to determine whether each document is relevant, instead of actually opening the document. As a result, the present invention facilitates database searching.

10 This brief summary has been provided so that the nature of the invention may be understood quickly. A more complete understanding of the invention can be obtained by reference to the following detailed description of the preferred embodiment thereof in connection with the attached drawings.

15

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a perspective view of a host computer system, including a personal computer and a query server, which is used in the present invention.

20

Figure 2 shows the architecture of the personal computer shown in Figure 1.

Figure 3 shows the architecture of the query server shown in Figure 1.

25 Figure 4 is a flow diagram showing process steps for setting-up the search hosting system of the present invention on a Web site.

Figure 5 shows the software architecture used to implement the process steps shown in Figure 4.

Figures 6A and 6B show a representative new account form used in setting-up the search hosting system of the present invention.

5 Figure 7 shows an example of a Web site which does not include the search hosting system of the present invention.

Figure 8 shows an example of a Web site which includes the search hosting system of the present invention.

10 Figure 9 shows the software architecture of the search hosting system of the present invention.

Figures 10A and 10B show process steps implemented by the software architecture of Figure 9 to perform a search of a Web site at a host computer system.

15 Figure 11 shows operation of a concordance generator included in the software architecture of Figure 9.

Figure 12 shows an example of a results page generated by the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

20 Figure 1 shows a representative embodiment of a host computer system which can be used to implement the present invention. As shown in Figure 1, host computer system 1 includes PC 2 and query server 4. PC 2 includes a local area network connection 5 for interfacing to query server 4, a general network connection 6 for interfacing to a
25 network such as the Internet, and fax/modem connection 7 for interfacing with other remote sources. PC 2 also includes display screen 9 for displaying information, keyboard 10 for inputting text and user commands, mouse 11 for positioning a cursor on display screen 9 and for inputting user commands, disk drive 12 for reading from and writing to floppy disks
30 installed therein, and CD-ROM drive 14 for accessing information stored on CD-ROM. PC 2 may also have one or more peripheral devices (not shown) attached thereto.

Figure 2 shows the internal structure of PC 2. As shown in

Figure 2, PC 2 includes memory 15, which comprises one or more computer-readable media, such as a computer hard disk. Memory 15 stores data 16, applications, and an operating system 17. Among the applications stored in memory 15 is results page generator 19, the function of which is described in detail below. Also included in PC 2 are display interface 20, keyboard interface 21, mouse interface 22, disk drive interface 24, CD-ROM drive interface 25, computer bus 26, RAM 27, and processor 29. Processor 29 preferably comprises a microprocessor or the like for executing applications out of RAM 27. As noted above, these applications may be stored in memory 15 or, alternatively, on a floppy disk in disk drive 12 or a CD-ROM in CD-ROM drive 14. In this regard, processor 29 accesses applications (or other data) stored on a floppy disk via disk drive interface 24 and accesses applications (or other data) stored on a CD-ROM via CD-ROM drive interface 25.

Application execution and other tasks of PC 2 and query server 4 may be controlled and/or altered using keyboard 10 or mouse 11, commands from which are transmitted to processor 29 via keyboard interface 21 and mouse interface 22, respectively. Output results from some applications running on PC 2 may be processed by display interface 20 and then displayed to a user on display 9. To this end, display interface 20 preferably comprises a display processor for forming images based on data provided by processor 29 over computer bus 26, and for outputting those images to display 9. Other applications, such as results page generator 19 receive their inputs from, and/or provide outputs to, network interfaces 5 and 6.

Figure 3 shows a representative architecture for query server 4. As shown in Figure 3, query server 4 includes processor 30, RAM 31, computer bus 32, and memory 34. Memory 34 preferably comprises one or more computer-readable media, such as a computer hard disk. Memory 34 stores data and applications, with the applications including Web robot 35, Web search engine 36, indexer module 37, and search and index hosting application 39, among others. Search and index hosting application 39 comprises computer-executable process steps to generate and execute a

database-specific search engine. A detailed description of these process steps is provided below.

5 In preferred embodiments of the invention, search engine 36 and indexer module 37 are based upon the search engine and indexer described in "The SMART Retrieval System: Experiments in Automatic Document Processing" by Gerald Salton (Prentice-Hall, Inc. (1971)) and "A Theory of Indexing" also by Gerald Salton (J. W. Arrowsmith, Ltd. (1975)). The contents of these two documents are hereby incorporated by reference into the subject application as if set forth herein in full. Web robot 35 can comprise any commercially available Web robot.

10 Applications such as search and index hosting application 39 are executed by processor 30 out of RAM 31, and results therefrom are output to host computer 2 via network connection 5. In this regard, search and index hosting application 39 comprises a database-specific search engine, in that it identifies which documents in a database contain a user-specified query term. The application then retrieves a list of such documents and, optionally, one or more data segments from each document on the list. In the preferred embodiment of the present invention, the database searched by application 39 is a Web site and the documents comprise one or more Web pages. However, it is noted that the invention is not limited to searching Web sites and, in fact, can be used to search a variety of different types of databases. Moreover, it is noted that although the invention will be described with respect to retrieving documents from Web sites based on text/character strings, the invention is not limited to this either. That is, the invention may also be used to retrieve images or other figures from any type of database.

Prior to performing site-specific searching using the present invention, it is necessary for a Web site provider to set up an account.

Figures 4 and 5 show how this is done. More specifically, as shown in Figure 5, a Web site provider's representative, such as their system administrator, issues request 40, over network 41 (e.g., the Internet), to a host computer system running search and index hosting application 39. In response, search and index hosting application 39 retrieves Web page forms required to set up a new account from storage 42, and transmits these forms back to the provider's representative, where the forms are displayed on the representative's computer. As shown in Figure 5, displayed form 44 requests "contact" information (e.g., information regarding the provider's Web site) and entry point URLs for the site. This contact information may be input in step S401 of Figure 4.

Figures 6A and 6B show a representative example of a new account form generated by a preferred embodiment of the present invention, which will be sold under the tradename PicoSearch™. As shown in these figures, new account form 45 requests information regarding the provider's representative, such as their name, E-mail address, and account password, together with three entry point URL's to the provider's Web site. Of course, the invention is not limited to using three entry point URLs, and even can be used without entering the contact information shown in Figure 6A. As shown in Figure 6B, the preferred embodiment of the invention also provides a way in which to control searching throughout a specified site. For example, in this embodiment, it is possible to restrict searching to the directory level, the server level, or the domain level. Of course, the invention is not limited to the restrictions shown in Figure 6B, and can be modified to include other searching options as well.

Once the contact information and entry point URLs have been added to the new account form, the "filled-out" form is transmitted over the network to the host computer system. There, contact information 46 and entry point URLs 48 are extracted from the new account form, as shown in Figure 5. The contact information (i.e., the provider representative's name, E-mail address, etc.) is then stored in user database

49, which may reside on memory 34 or any other memory on the host computer system. Search and index hosting application 39 then assigns the provider representative a new provider identifier 50 in step S402. This provider identifier is then combined with entry point URLs 48, the combination of which is provided to Web robot 35. At about the same time or shortly thereafter, Web site provider 52 provides data from its Web site (e.g., plural Web documents/pages) to host computer system 1 via network 41. In step S403, Web robot 35 searches through the data from the Web site for locations (i.e., URLs) and titles of documents therein, starting at the entry point URLs specified in the new account form. Next, in step S404, the location and title of each Web document that is found in the search is stored in user database 49, together with the provider identifier and a predetermined document identifier for each document. The use of these document identifiers is described below. After Web robot 35 has completed its "crawl" of the Web site, Web robot 35 passes provider identifier 50 for the Web site to indexer module 37.

In step S405, indexer module 37 indexes data stored in documents on the Web site. More specifically, indexer module 37 reviews each of Web documents 55 identified by Web robot 35, and creates an index of pointers to data in the Web site, where the data is indexed according to document, location in the document, and provider identifier. This index, together with the provider identifier and predetermined document identifiers, is then stored in multiple index storage 56 in step S406. In preferred embodiments of the invention, multiple index storage 56 resides in memory 34 of query server 4; although the invention is not limited to storing multiple index storage 56 there.

In addition to indexing data in the manner described above, in step S407, indexer module 37 compresses data from the Web site and stores that data in compressed document full text storage 57, together with the provider identifier for the site. As was the case above, storage 57 can reside in memory 34; although the invention is not limited to this. Next, in step S408, indexer 37 outputs processing results 59, including the provider identifier, to the provider's representative, as shown in Figure 5. At this

point, it is noted that although Figure 5 shows the provider's representative at a location which is different from that of the Web site provider, this need not be the case. That is, the provider's representative and the provider may also be at the same location.

5 Once an account has been established for the Web site, the provider's representative should copy code provided by the host computer system into their Web site. A representative example of such code is shown in the Appendix attached hereto. Among other things, this code permits the Web site to provide an interface to search and index hosting application 39 on the host computer system. As an illustration, Figure 7 shows an example of a Web site 60 which does not include the site-specific searching capabilities of the present invention, while Figure 8 shows that same Web site with the site-specific searching capabilities. As shown in Figure 8, the site with the searching capabilities includes entry line 61 and search button 62. To search for a text or the like at Web site 60, all that must be done is to input the text on the entry line, and click on the search button. The searching process of the present invention is described below with respect to the remaining figures.

20 First, however, it is noted that although the invention has been described with respect to only one Web site provider, the invention is not limited to use with only one provider. In fact, the invention is able to handle searches of any number of Web sites. That is, any number of Web site providers' representatives may set up accounts with the host computer system. As a result, the host computer system may store indices of data from plural databases (e.g., plural Web sites), locations of documents in those plural databases, and compressed text from the plural databases. However, because the invention also stores with each of these a provider identifier, the invention is able to select the appropriate data readily, and thus to perform site-specific searches. This feature of the invention is described more fully below.

30 In brief, as shown Figure 9, a user initiates a search at first network site 64 (e.g., the user's computer) for data in a remote database at second network site 65 (e.g., an Internet site or Web site server) and the

invention conducts the search at third network site 66 (e.g., the host computer system). At this point, it is noted that although Figure 9 shows the three network sites 64, 65 and 66 at three different locations (i.e., at three computer systems) on the network, these network sites may reside at the same network location or, alternatively, at only two different network locations. In any event, in accordance with the invention, first network site 64 receives a provider identifier associated with the database from second network site 65 and then outputs a query containing user-specified data and the provider identifier from first network site 64 to third network site 66. Third network site 66 then searches for the user-specified data in a database at the third network site using the provider identifier. This database at third network site 66 includes data that corresponds to data stored in the remote database at the second network site. The above process is described in more detail below.

More specifically, as shown in Figure 9, a user at first network site 64 initiates contact to Web site 65 via a Web browser (not shown), such as Netscape®. In response, Web site 65 automatically provides the user's computer with its provider identifier 67, together with Web page(s) 68 from the site. An example of such a page is shown in Figure 8. Thereafter, the user may enter a query 71 comprised, e.g., of user-specified data such as character strings, text, or the like, into the entry line provided with the Web page. The query and the provider identifier are then packetized and packet 72 is passed, via network 69 (e.g., the Internet), to the host computer (i.e., PC 2). The host computer receives this packet in step S1001 of Figure 10A, extracts the provider identifier and data from the query, and passes this information on to query server 4 via local network 70 (i.e., over local network connection 5 shown in Figure 1).

Thereafter, in step S1002, index retriever 74 uses provider identifier 67 to retrieve, from multiple index storage 56, an index of data that has the same provider identifier as the Web site. This index with the same provider identifier, labeled 75 in Figure 9, is then passed to search engine 36, along with query 71. In step S1003, search engine 36 searches through index 75 for data specified in query 71 in order to identify

documents that contain the specified data. The search engine then retrieves identifiers 76 (e.g., numbers) for these documents from the index.

Thereafter, processing proceeds to step S1004.

5 Step S1004 determines whether data segments in each
identified document should be output along with a list of identified
documents. In a case that step S1004 determines that no data segments
should be output, processing proceeds to step S1005 of Figure 10B. As
shown in Figure 9, this processing corresponds to dashed line 77, which
bypasses concordance generator 78 (described below). Thus, in step
10 S1005, search and index hosting application 39 extracts locations (e.g.,
URLs) and titles of documents from user database 49 that have identifiers
which match the documents identified in step S1003.

 The extracted document locations and titles 79 are then
output, via local network 70, to results page generator 19. In step S1006,
15 results page generator 19 generates a results page showing titles of, and
URL links to, documents on the Web site which contain the data specified
in the query. This results page 80 is then passed to the user's computer via
network 69, where it is displayed. In preferred embodiments of the
invention, this results page also includes an entry line, such as entry line 81
20 shown in Figure 9, from which the user may initiate a new search of the
Web site. If no new searches are to be performed, processing ends.
Otherwise, the foregoing process is repeated for the new search.

Returning to Figure 10A, in a case that data segments, such as text excerpts or the like, are to be output along with the list of documents, processing proceeds to step S1007. In this regard, as noted above, the present invention includes the capability of also outputting, from each document on the list, a text or data segment which contains the user-specified data. These segments comprise target data (i.e., the user-specified data) together with data adjacent to the target data. That is, assuming that the target data comprises a word which matches the user-specified data, the data segment for that word might comprise, e.g., five words to the left of the word, the word itself, and five words to the right of the word. Of course, the invention is not limited to using eleven-word data segments, meaning that any size data segment may be used. In fact, in preferred embodiments, the size and/or locations of the data segments relative to the user-specified data are settable by the user, and thus can be varied as desired.

Thus, following step S1004, in step S1007 concordance generator 78 searches the documents identified in step S1003 for locations of the user-specified data. Figure 11 is a close-up view of the operation of concordance generator 78. More specifically, as shown in Figure 11, in step S1007 concordance generator 78 performs a boolean search through documents 76 for data that matches data specified in query 71. When matching data is found, step S1008 generates pointers into the corresponding compressed data based on pointers to locations in the index. The generated pointers identify the user-specified data in the compressed data, together with a predetermined amount of additional adjacent data needed to make up a data segment.

Roughly concurrently, in step S1009, concordance generator 78 extracts the provider identifier for the current Web site from index 75, and selects and retrieves compressed data 82 for that Web site from full compressed text storage 57. Thereafter, in step S1010, concordance generator 78 decompresses selected data segments that correspond to the pointers generated in step S1008. In preferred embodiments of the invention, only compressed data corresponding to the selected data

segments is decompressed. As noted in Figure 11, a data segment may start m characters before the query term and end n characters after the query term - only this segment would be decompressed. Thereafter, arrays 84 of concordance strings (i.e., decompressed data segments) are output
5 from concordance generator 78, together with corresponding document identifiers specifying the document from which each string originated.

Arrays 84 of concordance strings are then output from query server 4 to results page generator 19 via local network 70. Results page generator 19 uses the decompressed data segments to formulate its results
10 page. In addition, results page generator 19 extracts document identifiers 85 corresponding to the decompressed data segments, and passes these document identifiers back to query server 4. In step S1011, query server 4 extracts locations (e.g., URLs) and titles of documents from user database 49 that match the document identifiers, and passes this information back to
15 results page generator 19. Results page generator 19 formulates a results page which includes document titles, data segments, URLs, and the original query term, and passes the results page, together with the provider identifier, back to the user in step S1012. The results page is then displayed on the user's computer. As was the case above, the results page
20 may include an entry line, from which the user may conduct another search of the Web site specified by the provider identifier. If no additional search is to be conducted, processing ends. Otherwise, the foregoing process is repeated for the new search.

Figure 12 is an example of a results page generated by the preferred embodiment of the present invention. Specifically, this results
25 page was generated from a search of the Web site shown in Figure 8, in which the user-specified character string was "labor and employment". As shown in Figure 12, the results page includes document titles 86a, 86b and 86c, and corresponding URLs 87a, 87b and 87c for each document title.
30 In addition, the results page shown in Figure 12 includes excerpts 88a, 88b and 88c from each document, which include the user-specified character string. As noted above, however, this feature of the invention is optional, and the results page need not include these excerpts. As also shown in

Figure 12, the results page includes entry line 89, whereby the user may conduct a new search of the Web site shown in Figure 8 for a new search term.

5 The present invention has been described with respect to a particular illustrative embodiment. It is to be understood that the invention is not limited to the above-described embodiment and modifications thereto, and that various changes and modifications may be made by those of ordinary skill in the art without departing from the spirit and scope of the appended claims.

APPENDIX

```
< <!-- Begin PicoSearch Code -->
< <P> < <FORM METHOD="POST"
5 ACTION="http://www.picosearch.com/cgi-bin/ts.pl">
  < <INPUT TYPE="HIDDEN" NAME="index" VALUE="236">
  < <TABLE BGCOLOR="WHITE" CELLSPACING=0
    CELLPADDING=0 BORDER=0> < <TR> < <TD>
  < <TABLE BGCOLOR="WHITE" CELLSPACING=2
10 CELLPADDING=0 BORDER=0>
    < <TR> < <TD> < <A HREF="http://www.picosearch.com">
      < <IMG BORDER="0"
        SRC="http://www.picosearch.com/picosmall.gif"
        ALT="PicoSearch"> < </A> < </TD>
15 < <TD> < <INPUT TYPE="TEXT" NAME="query" VALUE=""
      SIZE="20"> < </TD>
      < <TD> < <INPUT TYPE="SUBMIT"
        VALUE="Search"> < </TD> < </TR>
      < </TD> < </TR> < </TABLE> < </TABLE> < </FORM>
20 < <!-- End PicoSearch Code -->
```

WHAT IS CLAIMED IS:

1. A method of identifying which documents in a database contain user-specified data, the method comprising the steps of:
 - 5 storing, in memory, indices of data in plural databases, the data being indexed by document, location in the document, and a provider identifier;
 - receiving the user-specified data and a provider identifier which corresponds to one of the plural databases;
 - 10 retrieving, from memory, an index of data for a database that corresponds to the provider identifier; and
 - identifying documents in the retrieved index that contain the user-specified data.
- 15 2. A method according to Claim 1, further comprising the step of outputting identities of the documents identified in the identifying step.
- 20 3. A method according to Claim 1, wherein the database is located on the Internet.
4. A method according to Claim 1, further comprising the steps of:
 - 25 searching for locations of the user-specified data in the documents identified in the identifying step;
 - storing, in memory, compressed data which corresponds to data stored in each of the plural databases;
 - generating pointers into the compressed data based on the search conducted in the searching step;
 - 30 selecting compressed data from memory that corresponds to the provider identifier;

decompressing segments of the compressed data selected in the selecting step that correspond to the pointers generated in the generating step; and

outputting the decompressed segments.

5

5. A method according to Claim 4, wherein the pointers identify the user-specified data in the compressed data, together with a predetermined amount of adjacent data.

10

6. A method according to Claim 5, wherein the amount and/or location of the adjacent data relative to the user-specified data are variable.

15

7. A method according to Claim 4, wherein each segment of decompressed data includes a document identifier which identifies a document containing the segment; and

wherein the method further comprises the steps of:

storing, in memory, locations of documents in the plural databases, together with corresponding document identifiers;

20

extracting, from memory, locations of documents which contain the segments of decompressed data based on the document identifiers associated with the segments; and

outputting the locations of the documents which contain the segments of decompressed data.

25

8. A method according to Claim 6, wherein the locations of the documents comprise uniform resource locators.

30

9. A method of retrieving a list of documents in a database which include user-specified data, and of retrieving one or more data segments from each document on the list, the method comprising the steps of:

storing an index of documents from the database, the index

including pointers corresponding to data in the database, where the pointers define data segments having a predetermined size;

identifying one or more documents in the database that contain the user-specified data based on the index;

5 creating a list of one or more pointers corresponding to each occurrence of the user-specified data in the documents identified in the identifying step;

 extracting each data segment in the database that contains the user-specified data based on the list of pointers created in the creating step;
10 and

 outputting a list of the documents in the database that contain the user-specified data, together with the extracted data segments.

10. A method according to Claim 9, wherein the database is
15 located on the Internet.

11. A method according to Claim 10, wherein the list of documents comprises one or more uniform resource locators and/or uniform resource indicators.
20

12. A method according to Claim 9, wherein the identifying step comprises performing a boolean search of the index for documents which contain the user-specified data.

25 13. A method according to Claim 9, further comprising the step of storing, in memory, compressed data which corresponds to the data stored in the database; and

 wherein the extracting step comprises extracting each data segment from the compressed data and decompressing each extracted data
30 segment prior to output in the outputting step.

14. A method according to Claim 13, wherein each data segment includes a document identifier which identifies a document that

includes the data segment;

wherein the method further comprises storing, in memory, locations of documents in the database, together with associated document identifiers for the documents; and

5 wherein the outputting step generates the list of documents by comparing document identifiers from extracted data segments with document identifiers associated with locations of documents, extracting locations of documents corresponding to the extracted data segments based on the document identifiers, and including locations of the documents in the
10 list of documents.

15 15. A method of initiating a search at a first network site for user-specified data in a remote database at a second network site and of conducting the search at a third network site, the method comprising the steps of:

receiving, at the first network site, a provider identifier associated with the database from the second network site;

inputting the user-specified data at the first network site;

20 outputting the user-specified data and the provider identifier from the first network site to the third network site; and

25 searching for the user-specified data in a database at the third network site using the provider identifier, where the database at the third network site includes data that corresponds to data stored in the remote database at the second network site.

30 16. A method according to Claim 15, wherein the third network site includes a memory which stores indices of data in plural databases, the plural databases including the remote database at the second network site, and the data in each index being indexed by document, location in each database, and provider identifier; and

wherein the searching step comprises:

retrieving, from the memory, an index of data for a database that corresponds to the provider identifier; and

identifying documents in the index that contain the user-specified data.

5 17. A method according to Claim 16, wherein the database at the third network site is published on the Internet.

18. A method of configuring a computerized searching system so that the searching system can be used to search a database comprised of one or more documents, the method comprising the steps of:

10 inputting information identifying the database;
assigning a provider identifier to the database;
searching through the database using the information input in the inputting step in order to identify locations of documents in the database;

15 storing, in memory, the locations of the documents in the database, together with the provider identifier;
indexing the documents in the database;
storing, in memory, an index of the documents generated in the indexing step, together with the provider identifier;

20 storing, in memory, data corresponding to data in the database, together with the provider identifier; and
outputting the provider identifier to the database.

19. A method according to Claim 18, further comprising the steps of:

receiving user-specified data and the provider identifier for the database;

5 retrieving, from memory, the index of documents for the database based on the provider identifier; and

identifying documents in the index that contain the user-specified data.

10 20. A method according to Claim 19, wherein the data corresponding to data in the database is stored in memory in compressed format; and

wherein the method further comprises the steps of:

15 searching for the user-specified data in the documents identified in the identifying step;

generating pointers into the compressed data based on the search conducted in the searching step;

selecting compressed data from memory that corresponds to the provider identifier;

20 decompressing segments of the compressed data selected in the selecting step that correspond to the pointers generated in the generating step; and

outputting the decompressed segments of compressed data, together with a list of documents containing the decompressed segments.

25 21. An apparatus for identifying which documents in a database contain user-specified data, the apparatus comprising:

30 memory means which stores computer-executable process steps and indices of data in plural databases, where the data is indexed by document, location in the document, and provider identifier; and

a processor which executes the process steps so as (i) to receive the user-specified data and a provider identifier which corresponds to one of the plural databases, (ii) to retrieve, from memory, an index of

data for a database that corresponds to the provider identifier, and (iii) to identify documents in the retrieved index that contain the user-specified data.

5 22. An apparatus according to Claim 21, wherein the processor further executes process steps so as to output identities of the documents identified in the identifying step.

10 23. An apparatus according to Claim 21, wherein the database is located on the Internet.

15 24. An apparatus according to Claim 21, wherein the processor further executes process steps so as (i) to search for locations of the user-specified data in the documents identified in the identifying step, (ii) to store, in memory, compressed data which corresponds to data stored in each of the plural databases, (iii) to generate pointers into the compressed data based on the search conducted in the searching step, (iv) to select compressed data from memory that corresponds to the provider identifier, (v) to decompress segments of the compressed data selected in the selecting step that correspond to the pointers generated in the generating step, and (vi) to output the decompressed segments.

20 25. An apparatus according to Claim 24, wherein the pointers identify the user-specified data in the compressed data, together with a predetermined amount of data adjacent to the user-specified data.

26. An apparatus according to Claim 25, wherein the amount and/or location of the adjacent data relative to the user-specified data are variable.

5 27. An apparatus according to Claim 24, wherein each segment of decompressed data includes a document identifier which identifies a document containing the segment; and
 wherein the processor further executes process steps so as (i) to store, in memory, locations of documents in the plural databases,
10 together with corresponding document identifiers, (ii) to extract, from memory, locations of documents which contain the segments of decompressed data based on the document identifiers associated with the segments, and (iii) to output the locations of the documents which contain the segments of decompressed data.

15 28. An apparatus according to Claim 26, wherein the locations of the documents comprise uniform resource locators.

20 29. An apparatus for retrieving a list of documents in a database which include user-specified data, and for retrieving one or more data segments from each document on the list, the apparatus comprising:

 memory means which stores computer-executable process steps and an index of documents from the database, the index including pointers corresponding to data in the database, where the pointers define
25 data segments having a predetermined size; and

 a processor which executes the process steps so as (i) to identify one or more documents in the database that contain the user-specified data based on the index, (ii) to create a list of one or more pointers corresponding to each occurrence of the user-specified data in the
30 documents identified in the identifying step, (iii) to extract each data segment in the database that contains the user-specified data based on the list of pointers created in the creating step, and (iv) to output a list of the documents in the database that contain the user-specified data, together with

the extracted data segments.

30. An apparatus according to Claim 29, wherein the database is located on the Internet.

5

31. An apparatus according to Claim 30, wherein the list of documents comprises one or more uniform resource locators and/or uniform resource indicators.

10

32. An apparatus according to Claim 29, wherein the identifying step executed by the processor comprises performing a boolean search of the index for documents which contain the user-specified data.

15

33. An apparatus according to Claim 29, wherein the processor further executes process steps so as to store, in memory, compressed data which corresponds to data stored in the database; and wherein, in the extracting step, the processor extracts each data segment from the compressed data and decompresses each extracted data segment prior to output.

20

34. An apparatus according to Claim 33, wherein each data segment includes a document identifier which identifies a document that includes the data segment;

25

wherein the processor further executes process steps so as to store, in memory, locations of documents in the database, together with associated document identifiers for the documents; and

30

wherein, in the outputting step, the processor generates the list of documents by comparing document identifiers from extracted data segments with document identifiers associated with locations of documents, extracting locations of documents corresponding to the extracted data segments based on the document identifiers, and including locations of the documents in the list of documents.

35. A system for initiating a search at a first network site for user-specified data in a remote database at a second network site and for conducting the search at a third network site, the system comprising:

5 a first network site which executes process steps so as (i) to receive a provider identifier associated with the database from the second network site, (ii) to input the user-specified data, and (iii) to output the user-specified data and the provider identifier from the first network site to the third network site; and

10 a third network site which executes process steps so as to search for the user-specified data in a database at the third network site using the provider identifier, wherein the database at the third network site includes data that corresponds to data stored in the remote database at the second network site.

15 36. A system according to Claim 35, wherein the third network site comprises:

memory means which stores the process steps and indices of data in plural databases, the plural databases including the remote database at the second network site, and the data in each index being indexed by document, location in the database, and provider identifier; and

20 a processor which executes the process steps so as to search for the user-specified data by (i) retrieving, from memory, an index of data for a database based on the provider identifier, and (ii) identifying documents in the index that contain the user-specified data.

25 37. A system according to Claim 36, wherein the database at the third network site is published on the Internet.

30 38. An apparatus for configuring a computerized searching system so that the searching system can be used to search a database comprised of one or more documents, the apparatus comprising:

a memory which stores computer-executable process steps;
and

a processor which executes the process steps so as (i) to input information identifying the database, (ii) to assign a provider identifier to the database, (iii) to search through the database using the information input in the inputting step in order to identify locations of documents in the database, (iv) to store, in memory, the locations of the documents in the database, together with the provider identifier, (v) to index the documents in the database, (vi) to store, in memory, an index of the documents generated in the indexing step, together with the provider identifier, (vii) to store, in memory, data corresponding to data in the database, together with the provider identifier, and (viii) to output the provider identifier to the database.

39. An apparatus according to Claim 38, wherein the processor further executes process steps so as (i) to receive user-specified data and the provider identifier for the database, (ii) to retrieve, from memory, the index of documents for the database based on the provider identifier, and (iii) to identify documents in the index that contain the user-specified data.

40. An apparatus according to Claim 39, wherein the data corresponding to data in the database is stored in memory in compressed format; and

wherein the processor further executes process steps so as (i) to search for the user-specified data in the documents identified in the identifying step, (ii) to generate pointers into the compressed data based on the search conducted in the searching step, (iii) to select compressed data from memory that corresponds to the provider identifier, (iv) to decompress segments of the compressed data selected in the selecting step that correspond to the pointers generated in the generating step, and (v) to output the decompressed segments of data, together with a list of documents containing the decompressed segments.

41. Computer-executable process steps stored on a

computer-readable medium, the computer-executable process steps to identify which documents in a database contain user-specified data, the computer-executable process steps comprising:

5 code to store, in memory, indices of data in plural databases, the data being indexed by document, location in the document, and provider identifier;

 code to receive the user-specified data and a provider identifier which corresponds to one of the plural databases;

10 code to retrieve, from memory, an index of data for a database that corresponds to the provider identifier; and

 code to identify documents in the retrieved index that contain the user-specified data.

42. Computer-executable process steps according to Claim
15 41, further comprising code to output identities of the documents identified by the identifying code.

43. Computer-executable process steps according to Claim
20 41, wherein the database is located on the Internet.

44. Computer-executable process steps according to Claim
41, further comprising:

 code to search for locations of the user-specified data in the documents identified by the identifying code;

code to store, in memory, compressed data which corresponds to data stored in each of the plural databases;

code to generate pointers into the compressed data based on the search conducted by the searching code;

5 code to select compressed data from memory that corresponds to the provider identifier;

code to decompress segments of the compressed data selected by the selecting code that correspond to the pointers generated by the generating code; and

10 code to output the decompressed data segments.

45. Computer-executable process steps according to Claim 44, wherein the pointers identify the user-specified data in the compressed data, together with a predetermined amount of data adjacent to the user-specified data.

46. Computer-executable process steps according to Claim 45, wherein the amount and/or location of the data adjacent to the user-specified data are variable.

20 47. Computer-executable process steps according to Claim 44, wherein each segment of decompressed data includes a document identifier which identifies a document containing the segment; and

 wherein the computer-executable process steps further comprise:

25 code to store, in memory, locations of documents in the plural databases, together with corresponding document identifiers;

 code to extract, from memory, locations of documents which contain the segments of decompressed data based on the document identifiers associated with the segments; and

30

code to output the locations of the documents which contain the segments of decompressed data.

5 48. Computer-executable process steps according to Claim 46, wherein the locations of the documents comprise uniform resource locators.

10 49. Computer-executable process steps stored on a computer-readable medium, the computer-executable process steps to retrieve a list of documents in a database which include user-specified data, and to retrieve one or more data segments from each document on the list, the computer-executable process steps comprising:

15 code to store an index of documents from the database, the index including pointers corresponding to data in the database, where the pointers define data segments having a predetermined size;

 code to identify one or more documents in the database that contain the user-specified data based on the index;

20 code to create a list of one or more pointers corresponding to each occurrence of the user-specified data in the documents identified by the identifying code;

 code to extract each data segment in the database that contains the user-specified data based on the list of pointers created by the creating code; and

25 code to output a list of the documents in the database that contain the user-specified data, together with the extracted data segments.

 50. Computer-executable process steps according to Claim 49, wherein the database is located on the Internet.

51. Computer-executable process steps according to Claim 50, wherein the list of documents comprises one or more uniform resource locators and/or uniform resource indicators.

5 52. Computer-executable process steps according to Claim 49, wherein the identifying code comprises code to perform a boolean search of the index for documents which contain the user-specified data.

10 53. Computer-executable process steps according to Claim 49, further comprising code to store, in memory, compressed data which corresponds to data stored in the database;

wherein the extracting code comprises code to extract each data segment from the compressed data and to decompress each extracted data segment prior to output by the outputting code.

15 54. Computer-executable process steps according to Claim 53, wherein each data segment includes a document identifier which identifies a document that includes the data segment;

20 wherein the computer-executable process steps further comprise code to store, in memory, locations of documents in the database, together with associated document identifiers for the documents; and

25 wherein the outputting code comprises code to generate the list of documents by comparing document identifiers from extracted data segments with document identifiers associated with locations of documents, extracting locations of documents corresponding to the extracted data segments based on the document identifiers, and including locations of the documents in the list of documents.

55. Computer-executable process steps stored on a computer-readable medium, the computer-executable process steps to initiate a search at a first network site for user-specified data in a remote database at a second network site and to conduct the search at a third network site, the
5 computer-executable process steps comprising:

code to receive, at the first network site, a provider identifier associated with the database from the second network site;
code to input the user-specified data at the first network site;
code to output the user-specified data and the provider
10 identifier from the first network site to the third network site; and
code to search for the user-specified data in a database at the third network site using the provider identifier, where the database at the third network site includes data that corresponds to data stored in the remote database at the second network site.

15

56. Computer-executable process steps according to Claim 55, wherein the third network site includes a memory which stores indices of data in plural databases, the plural databases including the remote database at the second network site, and the data in each index being
20 indexed by document, location in the database, and provider identifier; and

wherein the searching code comprises:
code to retrieve, from memory, an index of data for a database that corresponds to the provider identifier; and
code to identify documents in the index that contain the user-
25 specified data.

57. Computer-executable process steps according to Claim 56, wherein the database at the third network site is published on the Internet.

58. Computer-executable process steps stored on a computer-readable medium, the computer-executable process steps to configure a computerized searching system so that the searching system can be used to search a database comprised of one or more documents, the computer-executable process steps comprising:

code to input information identifying the database;
code to assign a provider identifier to the database;
code to search through the database using the information input by the inputting code in order to identify locations of documents in the database;
code to store, in memory, the locations of the documents in the database, together with the provider identifier;
code to index the documents in the database;
code to store, in memory, an index of the documents generated in the indexing step, together with the provider identifier;
code to store, in memory, data corresponding to data in the database, together with the provider identifier; and
code to output the provider identifier to the database.

59. Computer-executable process steps according to Claim 58, further comprising:

code to receive user-specified data and the provider identifier for the database;
code to retrieve, from memory, the index of documents for the database based on the provider identifier; and
code to identify documents in the index that contain the user-specified data.

60. Computer-executable process steps according to Claim 59, wherein the data corresponding to data in the database is stored in memory in compressed format; and

5 wherein the computer-executable process steps further comprise:

 code to search for the user-specified data in the documents identified by the identifying code;

 code to generate pointers into the compressed data based on the search conducted by the searching code;

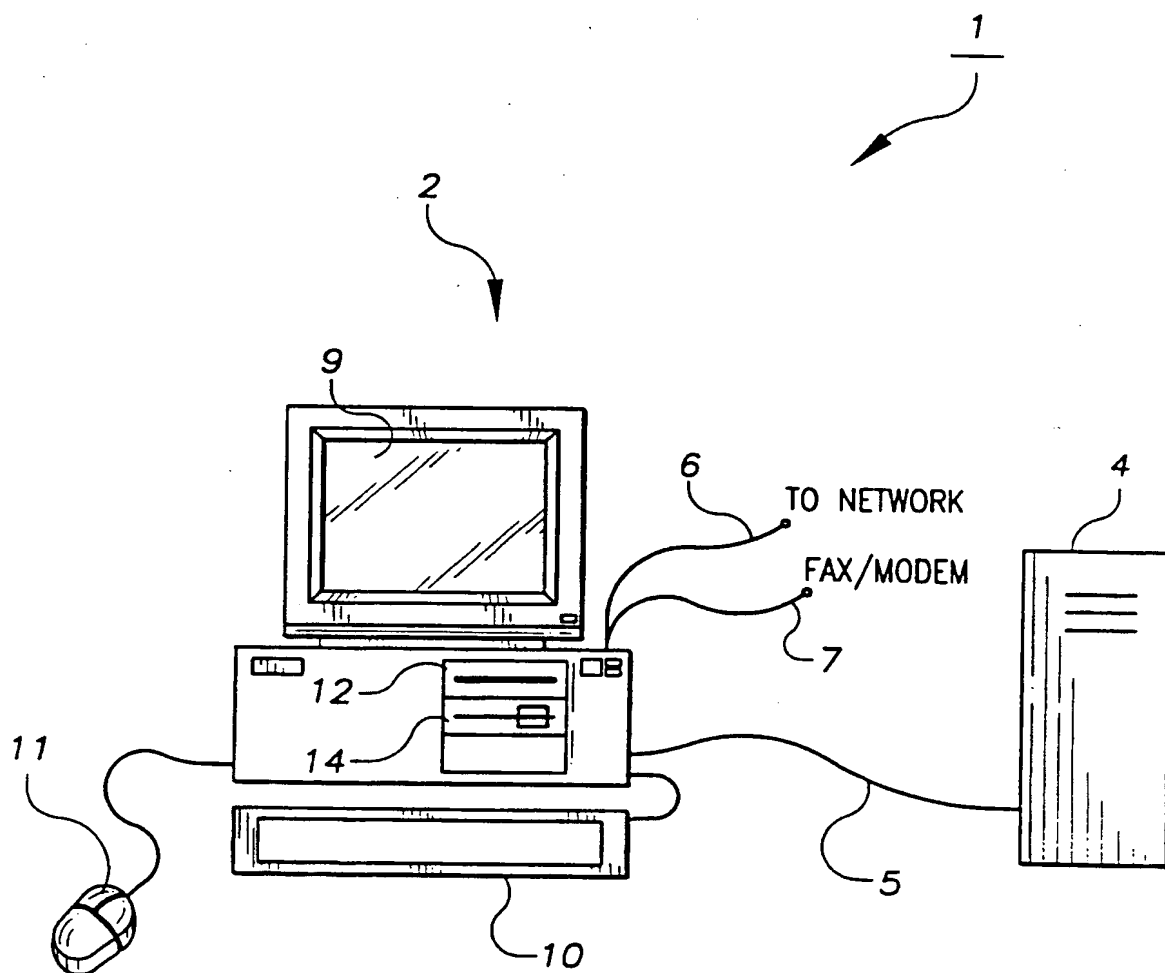
10 code to select compressed data from memory that corresponds to the provider identifier; -

 code to decompress segments of the compressed data selected by the selecting code that correspond to the pointers generated by the generating code; and

15 code to output the decompressed segments of compressed data, together with a list of documents containing the decompressed segments.

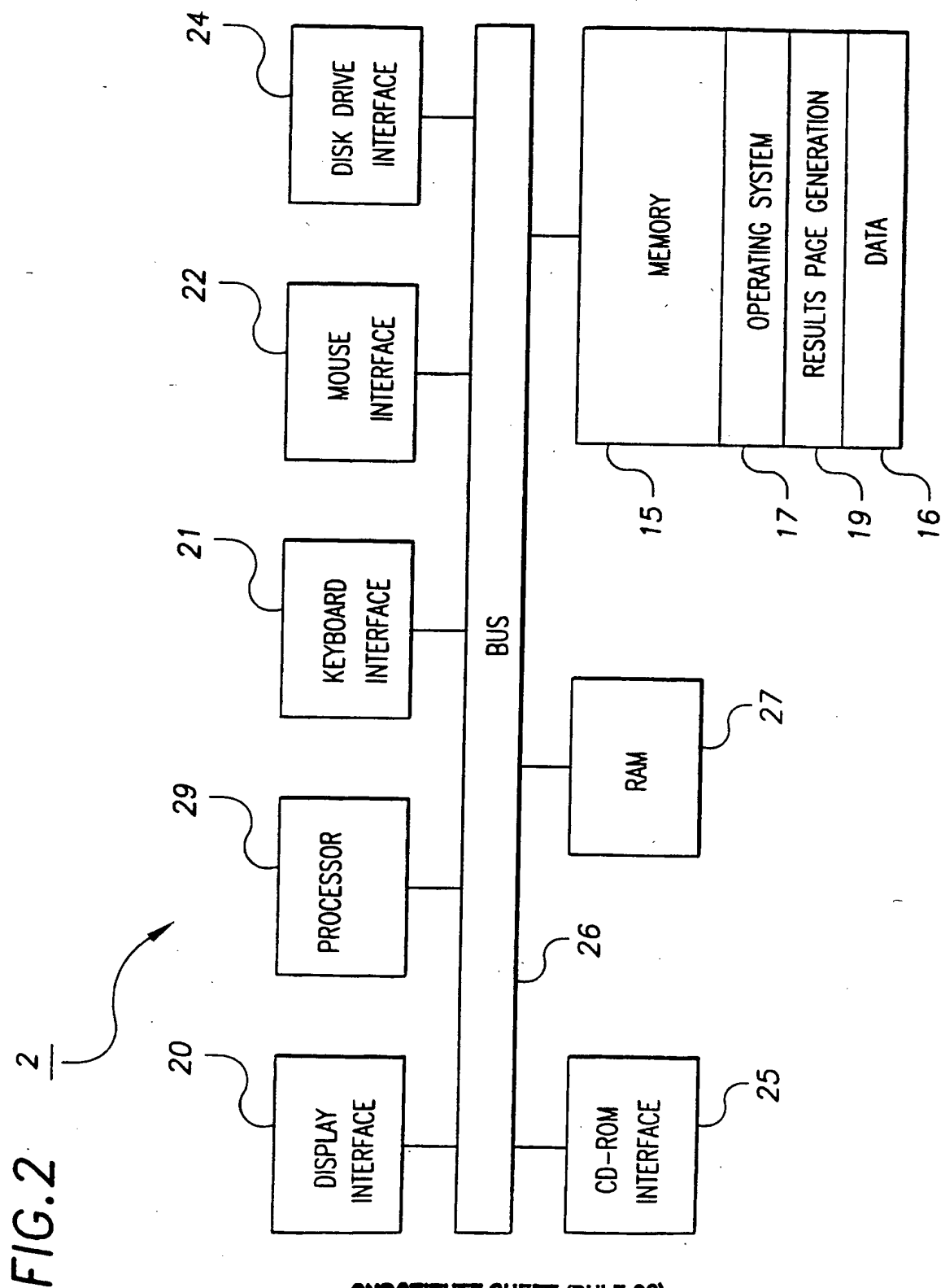
1/18

FIG. 1



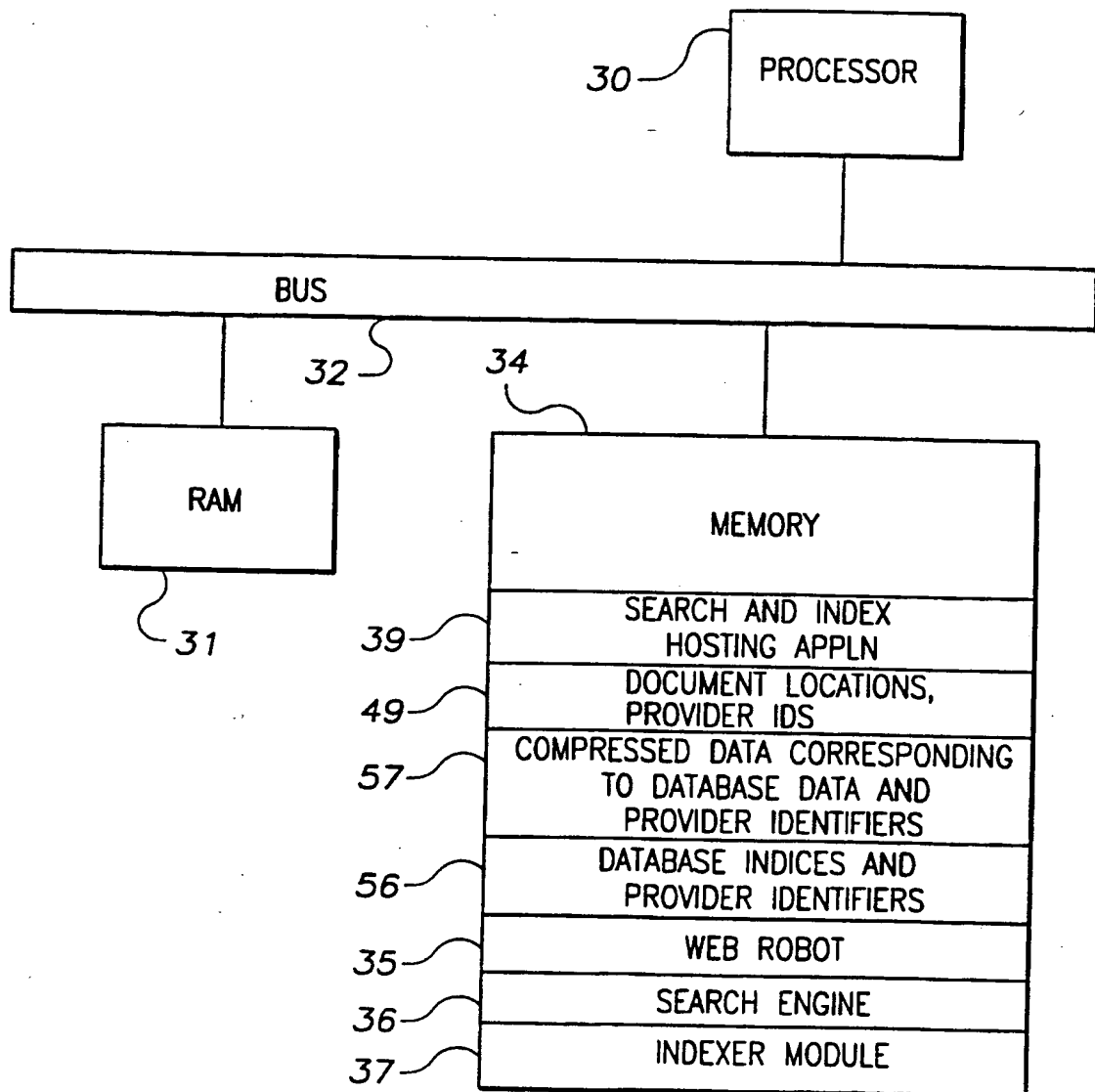
SUBSTITUTE SHEET (RULE 26)

2/18



3/18

FIG. 3



SUBSTITUTE SHEET (RULE 26)

4/18

FIG. 4

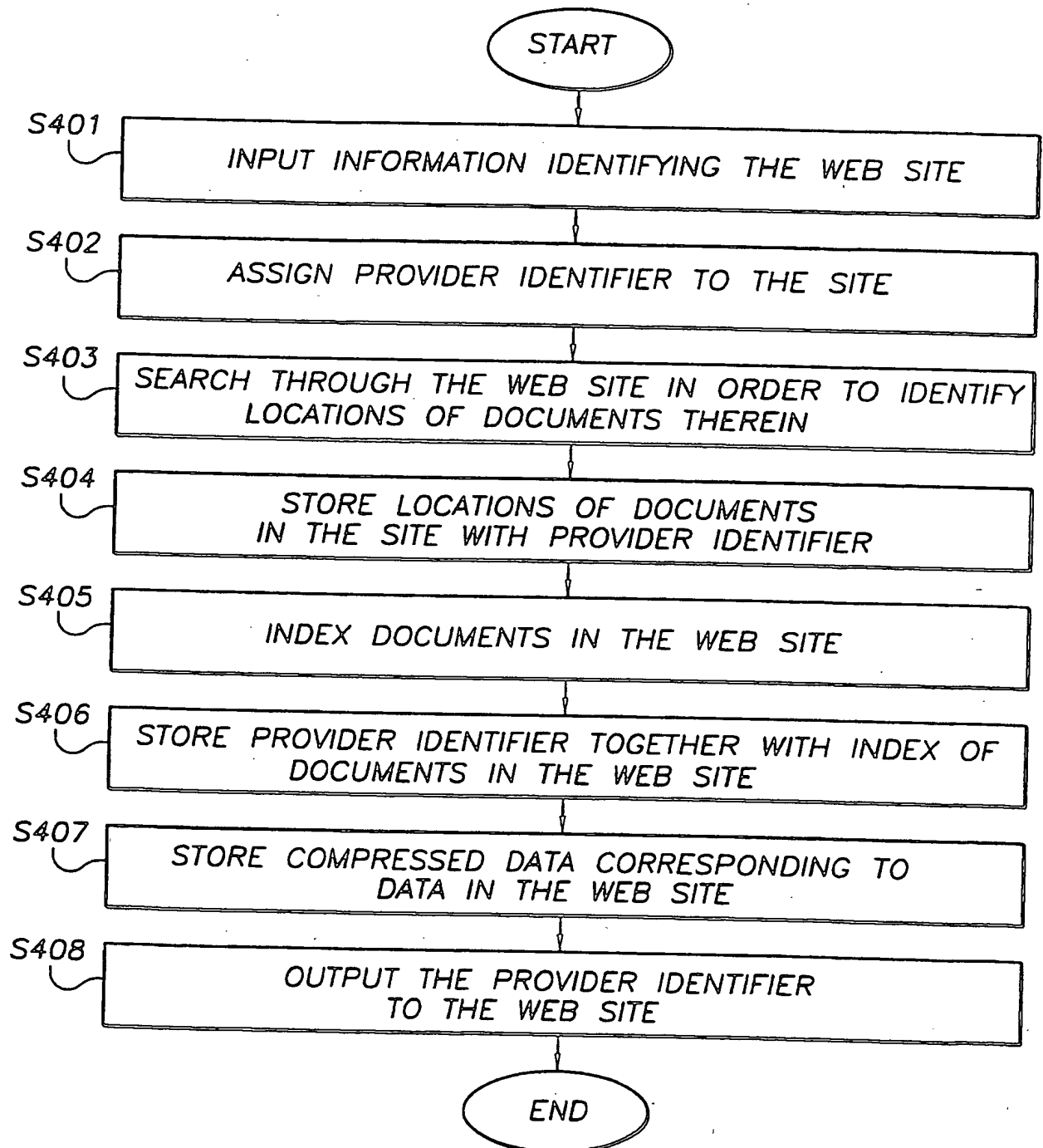
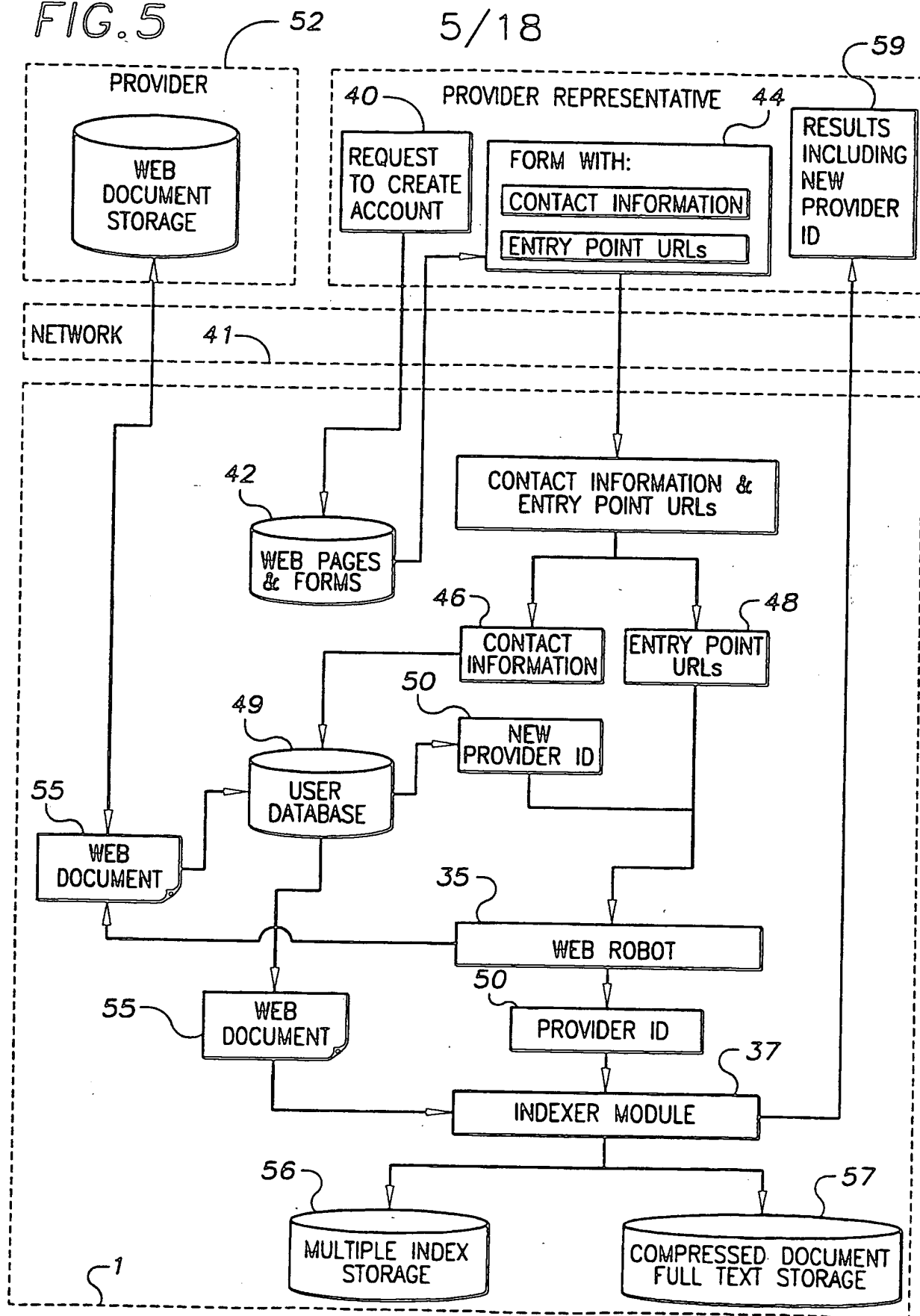


FIG. 5

5/18



SUBSTITUTE SHEET (RULE 26)

6/18


FIG. 6A-1


45


pico Search™
SEARCH HOSTING™


NEW ACCOUNT


PLEASE PROVIDE US WITH THE FOLLOWING INFORMATION ABOUT YOU AND YOUR WEB SITE SO THAT WE CAN BEGIN BUILDING YOUR PicoSearch SEARCH ENGINE.



NAME


CURRENT
USERS


TRY NOW
FOR FREE


TECHNICAL
SPACE


FAQS


PLAN \$

- LITE (FREE)
- PERSONAL
- BUSINESS

FIRST NAME

LAST NAME

E-MAIL ADDRESS

PLEASE PICK A PASSWORD:

PLEASE RE-TYPE YOUR PASSWORD:

NOW WE NEED YOU TO ENTER ALL OF THE ENTRY POINTS TO YOUR WEB SITE. FOR MOST PEOPLE, YOU WILL ONLY NEED TO ENTER YOUR MAIN ADDRESS, IN THE FORMAT OF:

TO FIG. 6A-2

7/18

FIG. 6A-2

FROM FIG. 6A-1

http://www.mysite.com/index.html

or

http://www.internet.com/~mypage/index.html

WE ARE GOING TO SEND OUR WEB ROBOT TO VISIT ALL YOUR PAGES SO THAT WE CAN MAKE A SEARCH INDEX ON THEM. IF YOU HAVE PAGES THAT ARE NOT ACCESSABLE BY

FOLLOWING LINKS FROM YOUR FIRST ENTRY POINT URL,

PLEASE ENTER THAT ADDRESS AS WELL. FOR EXAMPLE, YOU MAY HAVE TWO DIFFERENT SETS OF PAGES ON YOUR SITE THAT ARE NOT LINKED TO EACH OTHER AT ALL. YOU SIMPLY NEED TO ENTER THE MAIN URL FOR EACH SITE. FOR

EXAMPLE:

http://www.mysite.com/widgets/index.html

and

http://www.mysite.com/hammers/index.html

YOU MAY ENTER UP TO THREE (3) ENTRY POINTS TO YOUR SITE, IN THE FIELDS BELOW:

ENTRY POINT

#1

ENTRY POINT

#2

ENTRY POINT

#3

45

8/18

FIG. 6B-1

45

PLEASE TELL US HOW YOU WOULD LIKE TO LIMIT YOUR SEARCH. THERE ARE THREE OPTIONS:

☐ DIRECTORY LEVEL RESTRICTION

IF YOU SPECIFY YOUR FIRST ENTRY POINT WITH THE FULL FILE NAME (FOR EXAMPLE:

http://www.mysite.com/-mydirectory/index.html)

THEN WE CAN LIMIT OUR SEARCH AND INDEXING TO ONLY THAT DIRECTORY.

TO FIG. 6B-2

9/18

FIG. 6B-2

FROM FIG. 6B-1

☒ SERVER LEVEL RESTRICTION (DEFAULT)
 YOUR SEARCH WILL BE LIMITED TO THE WEB SERVER THAT
 YOU SPECIFY IN YOUR FIRST ENTRY POINT. SO IF YOU
 SPECIFY: *http://www.mysite.com/sompage.html* THEN
 WE WILL LOOK FOR OTHER PAGES ON *www.mysite.com* BUT
 NOT ON OTHER SERVERS IN YOUR DOMAIN, LIKE
cgi.mysite.com.

☐ DOMAIN LEVEL RESTRICTION
 YOUR SEARCH WILL BE LIMITED ONLY TO YOUR DOMAIN.
 SO, IF YOU TYPE IN
http://www.mysite.com/sompage.html AS YOUR FIRST
 ENTRY POINT, AND WE FIND A LINK TO A PAGE ON
http://server2.mysite.com WE WILL FOLLOW THE LINK
 AND INDEX THOSE PAGES AS WELL.

PLEASE
 REVIEW YOUR
 SELECTIONS
 AND THEN
 CLICK:

NEXT

45

10/18

FIG. 7

60

Choate, Hall & Stewartwww.choate.com*A Partnership Including Professional Corporations*About the Firm | News & Updates | Firm Directory | Recruiting | Announcements | Mail*Exchange Place**53 State Street**Boston, Massachusetts 02109-2891**Telephone 617/248-5000**Fax 617/248-4000**E-Mail: info@choate.com*

Copyright 1996 by Choate, Hall & Stewart

Last Updated February 20, 1998

Parts of this Web Site may be considered advertising
under the rules of the Massachusetts Supreme Judicial Court

11/18

FIG. 8

60

Choate, Hall & Stewart
www.choate.com

A Partnership Including Professional Corporations

[About the Firm](#)

[News & Updates](#)

[Firm Directory](#)

[Recruiting](#)

[Announcements](#)

[Mail](#)

ptco Search
SEARCH HOSTING

62

Exchange Place 61
53 State Street
Boston, Massachusetts 02109-2891
Telephone 617/248-5000
Fax 617/248-4000
E-Mail: info@choate.com

Copyright 1996 by Choate, Hall & Stewart
Last Updated February 20, 1998

Parts of this Web Site may be considered advertising
under the rules of the Massachusetts Supreme Judicial Court

FIG. 9

12/18

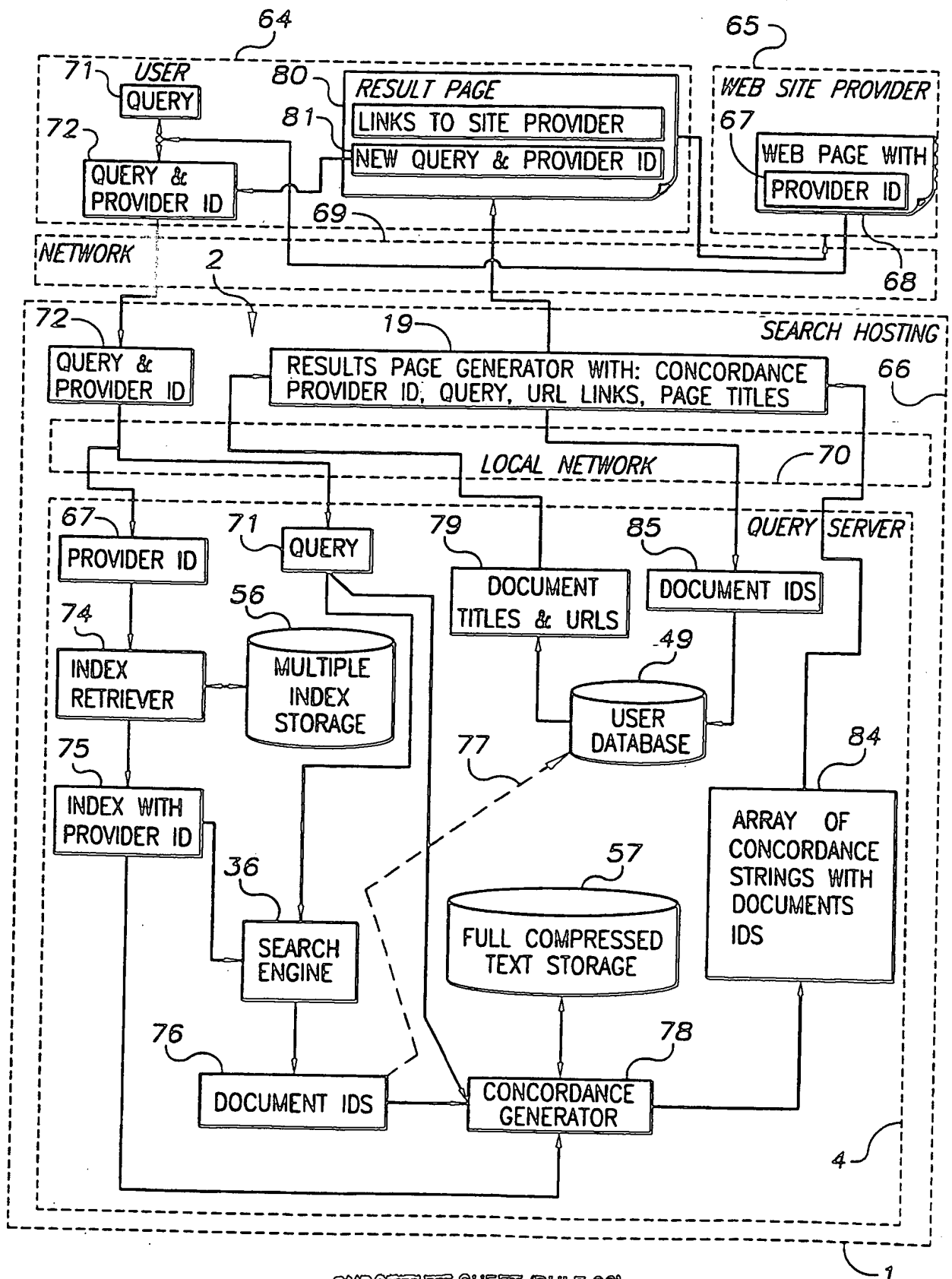
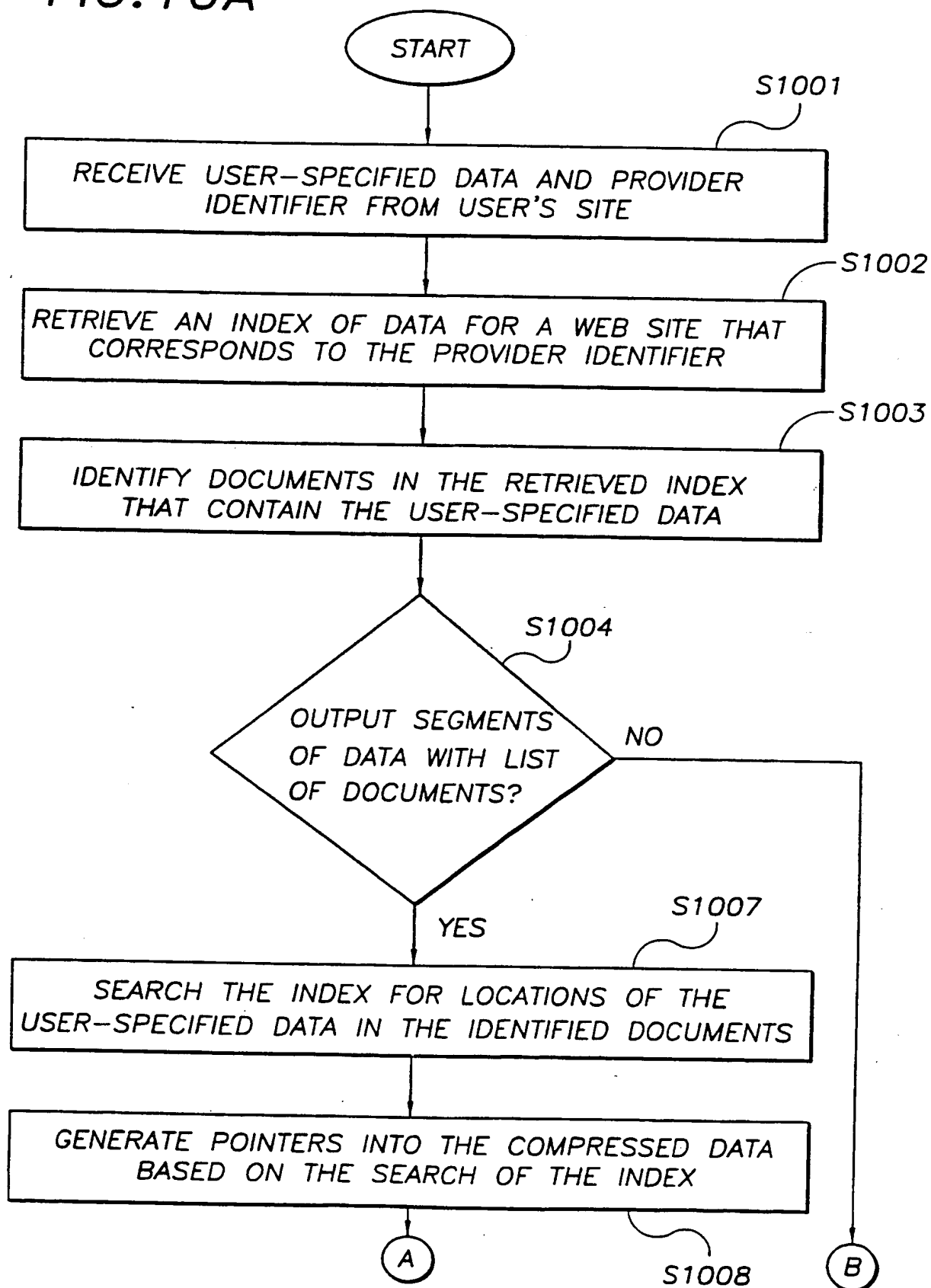


FIG. 10A

13/18



14/18

FIG. 10B

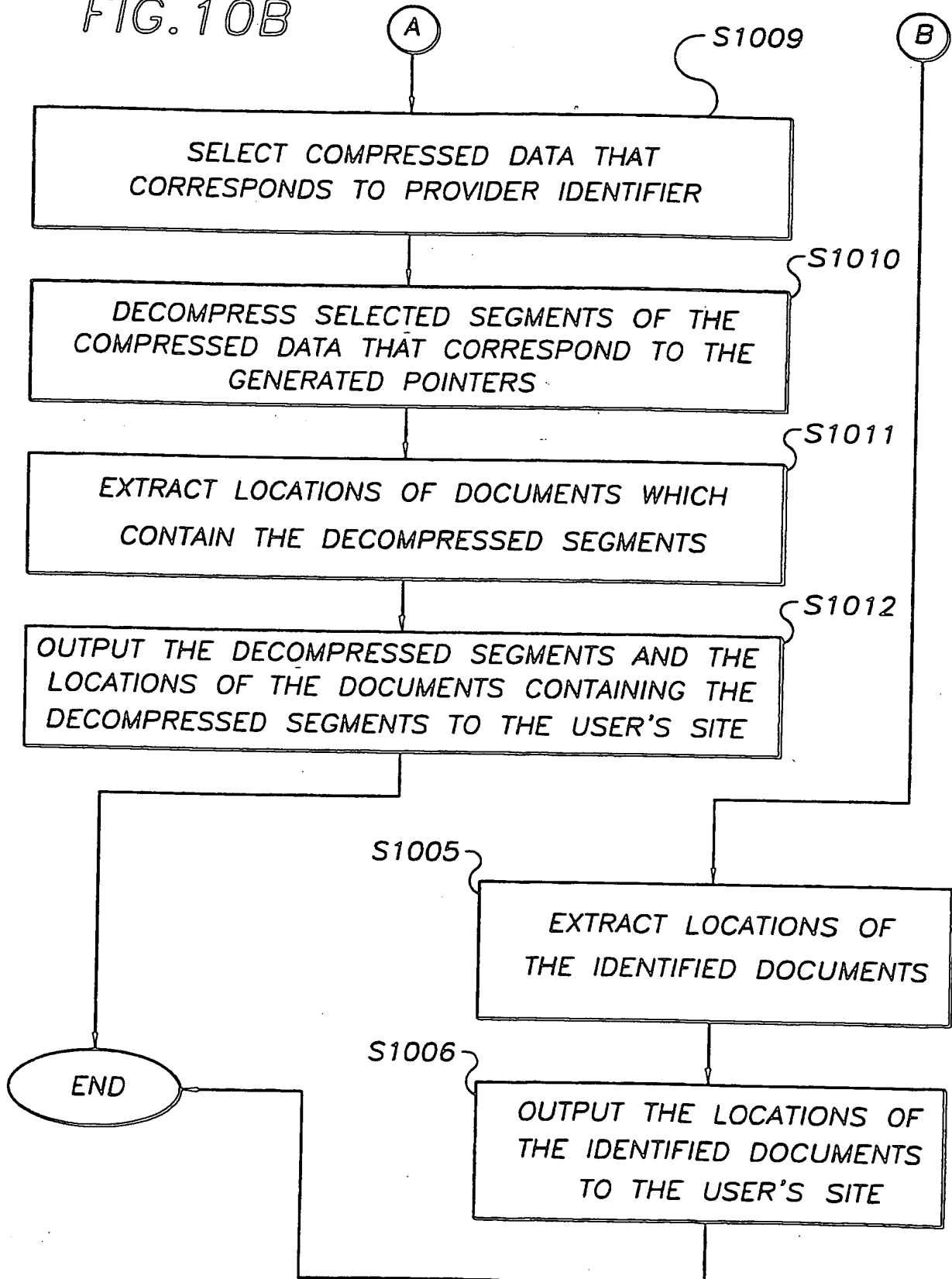
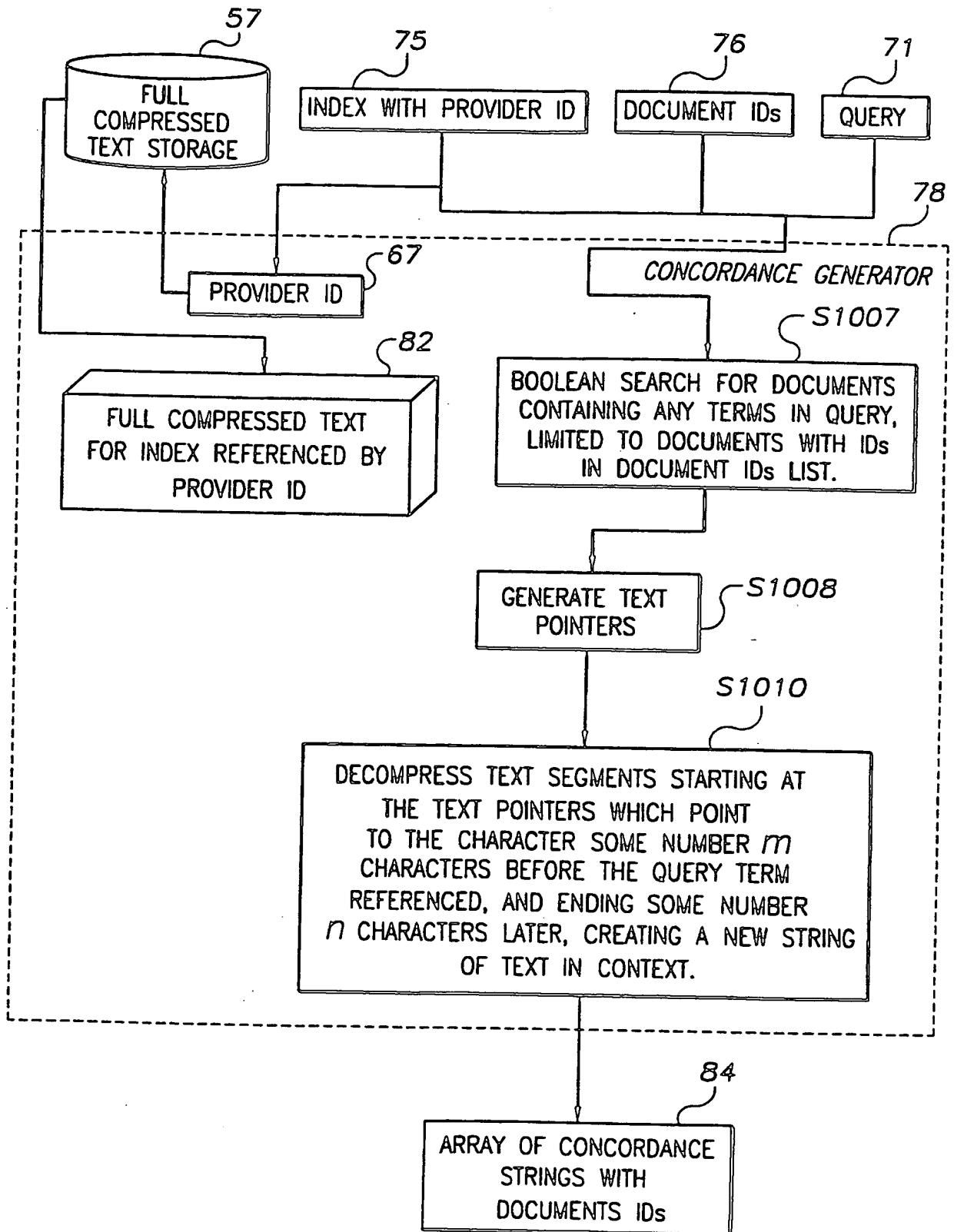


FIG. 11

15/18



16/18

FIG. 12A

PICO SEARCH
TERAGRAM
CORPORATION

Pico Search™ SEARCH HOSTING™ "LABOR AND EMPLOYMENT" SEARCH ⁸⁹

FOUND 3 DOCUMENTS FOR YOUR QUERY. ^{87a}

1. archv.html

^{86a} http://www.choate.com/News/Archives/archv.html
CHOATE, HALL & STEWART ARCHIVES ARTICLES/NEWSLETTERS
THE FOLLOWING IS A LIST OF ARTICLES.

LAST UPDATED: 7/3/98

^{88a} { ... 2 OCTOBER 1996 LABOR AND EMPLOYMENT LAW REPORT NEWS ON ...
... ALL AMP STEWARTS LABOR AND EMPLOYMENT GROUP VOLUME VN...
... UP JANUARY 1996 LABOR AND EMPLOYMENT LAW REPORT NEWS AND UPDATES...
... BER 1996 LABOR AND EMPLOYMENT LAW REPORT NEWS AND UPDATES...
... STEWART'S LABOR AND EMPLOYMENT GROUP VOLUME V NO 1 MAY...
... RY 1996 LABOR AND EMPLOYMENT LAW REPORT NEWS AND UPDATES...

TO FIG. 12B

17/18

FIG.12B

FROM FIG.12A

2. WHO WE ARE

86b { <http://www.choate.com/Recruiting/Who We Are.html> } 87b
 CHOATE, HALL & STEWART WHO WE ARE AMONG LAW FIRMS
 OF NATIONAL STANDING AND HIGH CAREER
 LAST UPDATED: 7/3/98

88b { ...ELLECTUAL PROPERTY LABOR AND EMPLOYMENT BANKRUPTCY SPORTS...
 ...PROPERTY LABOR AND EMPLOYMENT BANKRUPTCY SPORTS LENDING A ...

3. OUR PRACTICE

86c { <http://www.choate.com/Recruiting/OUR PRACTICE.html> } 87c
 CHOATE, HALL & STEWART OUR PRACTICE CHOATE, HALL &
 STEWART SERVES CORPORATE,

TO FIG.12C

18/18

FIG. 12C

FROM FIG. 12B

LAST UPDATED: 7/3/98

...ELLECTUAL PROPERTY LABOR AND EMPLOYMENT CONSTRUCTION AND Z...
 ...GENCIES AND COURTS LABOR AND EMPLOYMENT MATTERS INCLUDING...
 ...SITIONS INSURANCE LABOR AND EMPLOYMENT PATIENT RIGHTS RE ...
 ...UART CONSULTING ON LABOR AND EMPLOYMENT MATTERS AND DEVELOP
 ...
 ...PROPERTY LABOR AND EMPLOYMENT CONSTRUCTION AND ZONING LAN...
 ...D COURTS LABOR AND EMPLOYMENT MATTERS INCLUDING COLLECTIVE...
 ...NSURANCE LABOR AND EMPLOYMENT PATIENT RIGHTS REIMBURSEMEN...
 ...ULTING ON LABOR AND EMPLOYMENT MATTERS AND DEVELOPING AN EMP
 ...

88c

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/17359

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 98 07105 A (INFOSEEK CORP) 19 February 1998 (1998-02-19) page 1, line 1 -page 11, line 15	1-4, 9-12, 15, 18, 21, 29, 35, 38, 41, 49, 55, 58
A	K. RAILSBACK: "Compass Serever 3.0" ZDNET REVIEWS, 'Online! 13 February 1998 (1998-02-13), XP002126467 Retrieved from the Internet: <URL:http://www.zdnet.com/products/stories /reviews/0,4161,280094,00.html> 'retrieved on 1999-12-21! the whole document -/--	1, 9, 15, 18, 21, 29, 35, 38, 41, 49, 55, 58

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

21 December 1999

Date of mailing of the international search report

13/01/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Fournier, C

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/17359

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KOSTER M: "ALIWEB - Archie-like indexing in the WEB" COMPUTER NETWORKS AND ISDN SYSTEMS, NL, NORTH HOLLAND PUBLISHING. AMSTERDAM, vol. 27, no. 2, page 175-182 XP004037988 ISSN: 0169-7552 the whole document	1,9,15, 18,21, 29,35, 38,41, 49,55,58
A	US 5 745 890 A (BURROWS MICHAEL) 28 April 1998 (1998-04-28) abstract	1,4
A	EP 0 530 993 A (XEROX CORP) 10 March 1993 (1993-03-10) abstract	1,4,5

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/17359

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9807105	A	19-02-1998	US 5920854 A	06-07-1999
			AU 3485697 A	06-03-1998
US 5745890	A	28-04-1998	NONE	
EP 0530993	A	10-03-1993	US 5278980 A	11-01-1994
			DE 69229204 D	24-06-1999
			DE 69229204 T	04-11-1999
			JP 5189487 A	30-07-1993